A Summary of Markov Chain Monte Carlo(MCMC) Linkage Analysis

Jing Hua Zhao (j.zhao@iop.kcl.ac.uk)

July 3, 2001

The major references are Tanner (1996), Schafer (1997) and Thompson (1995, 2000a, 2000b, 2000c). Relevant references also include Heath (1997), Lin (1996, 1999) and Sheehan (2000).

1 EM, data augmentation, and MCMC

These methods are mostly encountered in context of missing data problem or Bayesian parameter estimation. For missing data problem a complete data is represented as $Y = (Y_{obs}, Y_{mis})$, Assuming that the complete-data model and ignorability assumptions are correct, all relevant statistical information about the parameters is contained in the observed-data likelihood $L(\theta|Y_{obs})$ or observed-data posterior $P(\theta|Y_{obs})$. Both EM (Dempster et al, 1977) and **data augmentation** (DA, Tanner and Wong 1987) solve a difficult problem by repeatedly solving tractable complete-data problem.

EM capitalises on the interdependence between missing data Y_{mis} and parameter θ. The fact that Y_{mis} contains information relevant to estimating θ, and θ in turn helps us to find likely values of Y_{mis}, suggests the following scheme for estimating θ in the presence of Y_{obs} alone:"Fill in" the missing data Y_{mis} based on an initial estimate of θ, re-estimate θ based Y_{obs} and the filled-in Y_{mis} and iterate until the estimates converges.

In any incomplete-data problem, the distribution of the complete data Y can be factored as

$$P(Y|\theta) = P(Y_{obs}|\theta)P(Y_{mis}|Y_{obs}, \theta).$$

Viewing each term as a function of θ, it follows that

$$l(\theta|Y) = l(\theta|Y_{obs}) + \log P(Y_{mis}|Y_{obs}, \theta) + c$$

where $l(\theta|Y)$ and $l(\theta|Y_{obs})$ are the complete-data and observed data log-likelihoods, respectively. The term $P(Y_{mis}|Y_{obs}, \theta)$ is the predictive distribution of the missing data given θ and c is a constant. Since Y_{mis} is unknown we can take average of this formulation over the predictive distribution $P(Y_{mis}|Y_{obs}, \theta^{(t)})$, where $\theta^{(t)}$ is a preliminary estimate of the unknown parameter. This yields

$$Q(\theta|\theta^{(t)}) = l(\theta|Y_{obs}) + H(\theta|\theta^{(t)}) + c,$$

where

$$Q(\theta|\theta^{(t)}) = \int l(\theta|Y)P(Y_{mis}|Y_{obs}, \theta^{(t)})dY_{mis}$$

and

$$H(\theta|\theta^{(t)}) = \int \log(Y_{mis}|Y_{obs},\theta)P(Y_{mis}|Y_{obs},\theta^{(t)})dY_{mis}.$$

A central result about EM is that if we let $\theta^{(t+1)}$ be the value of θ that maximises $Q(\theta|\theta^{(t)})$, then $\theta^{(t+1)}$ is better estimate than $\theta^{(t)}$ in the sense that its observed-data log-likelihood is at least as high as that of $\theta^{(t)}$.

$$l(\theta^{(t+1)}|Y_{obs}) \geq l(\theta^{(t)}|Y_{obs}).$$

This can be seen by writing out $l(\theta^{(t+1)}|Y_{obs}) - l(\theta^{(t)}|Y_{obs})$, the Q terms are non-negative whereas the H terms $\int \log\left[\frac{P(Y_{mis}|Y_{obs},\theta^{(t)})}{P(Y_{mis}|Y_{obs},\theta^{(t+1)})}\right] P(Y_{mis}|Y_{obs},\theta^{(t)})dY_{mis}$ is non-negative by Jensen's inequality and the convexity of the function $x\log x$.

Each EM iteration consists of two distinct steps:

1. The Expectation or E-step, in which the function $Q(\theta|\theta^{(t)})$ is calculated by averaging the complete-data log-likelihood $l(\theta|Y)$ over $P(Y_{mis}|Y_{obs},\theta^{(t)})$; and

2. The Maximisation or M-step, in which $\theta^{(t+1)}$ is found by maximising $Q(\theta|\theta^{(t)})$.

Alternately performing the E- and M- steps beginning with a starting value $\theta^{(0)}$ defines a sequence of iterates. Dempster et al. (1977) and Wu (1983) provide conditions under which this sequence converges reliably to a stationary point of the of the observed-data log-likelihood. In well-behaved problems this stationary point is a global maximum and EM yields the MLE of θ.

The ECM (Expectation-Conditional Maximisation, Meng and Rubin 1993) algorithm is a useful extension of the EM for situations where the M-step cannot be carried out without iteration. ECM replaces a complicated M-step with a simpler conditional or constrained maximisations known as CM-step. ECM retains the reliable convergence properties of EM while simplifying, and often reducing, the required computations. The CM step is comprised of S conditional maximisations in which the Q function is maximised not over the entire parameter space but over a smaller set in which a vector-valued function $g_s(\theta)$ is fixed at its previous value for $s = 1, 2, \ldots, S$. The set of functions must be pre-selected and satisfy certain precise conditions. Given the current value of the parameter $\theta^{(t)}$, E-step is applied to obtain $Q(\theta|\theta^{(t)})$ as in EM algorithm, then obtain $\theta^{(t+1)}$ by maximising $Q(\theta|\theta^{(t)})$ subject to constraints $g_s(\theta) = g_s(\theta^{(t+(s-1)/S)})$, $s = 1, 2, \ldots, S$. The resulting parameter value $\theta^{(t+S/S)} = \theta^{(t+1)}$ becomes the input to the next E-step.

Suppose that a random vector z is partitioned into two subvectors, $z = (u, v)$, where the joint distribution $P(z)$ is not easily simulated but the conditional distributions $P(u|v) = g(u|v)$ and $P(v|u) = h(v|u)$ are. DA was motivated from the representation of the posterior density:

$$p(\theta|y) = \int_z p(\theta|z,y)p(z|y)dz$$

$p(\theta|y)$ denotes the posterior density of the parameter θ given the data y. $p(y|z)$ denotes the predictive density of latent data z given y, and $p(\theta|y,z)$ denotes the conditional density of θ given the augmented data (y,z).

Similarly

$$p(z|y) = \int_\Theta p(z|\phi,y)p(\phi|y)d\phi$$

and by substitution $p(\theta|y)$ satisfy

$$g(\theta) = \int K(\theta,\phi)g(\phi)d\phi, \quad K(\theta,\phi) = \int p(\theta|z,y)p(z|\phi,y)dz$$

Since

$$\begin{aligned} p(\theta|y) &= \int_z p(\theta|z,y)\int_\Theta p(z|\phi,y)p(\phi|y)d\phi dz \\ &= \int_\Theta p(\theta|z,y)p(z|\phi,y)dz \int_z p(\phi|y)d\phi \end{aligned}$$

If there exists a operator T that $Tf(\theta) = \int K(\theta,\phi)f(\phi)d\phi$, it hints $g_{i+1}(\theta) = (Tg_i)(\theta)$ and a two-step algorithm (a) obtain $z^{(1)},...,z^{(m)} \sim p(z|y)$, (b) update $p(\theta|y)$: $g_{i+1}(\theta) = m^{-1}\sum_{j=1}^m p(\theta|z^{(j)},y)$. The data augmentation satisfies these.

Consider the following iterative sampling scheme: given a current guess $\theta^{(t)}$ of the parameter, first draw a value of the missing data from the conditional predictive distribution of Y_{mis},

$$Y_{mis} \sim P(Y_{mis}|Y_{obs},\theta^{(t)}).$$

Then conditioning on $Y_{mis}^{(t+1)}$, draw a new value of θ from its complete-data posterior,

$$\theta^{(t+1)} \sim P(\theta|Y_{obs},Y_{mis}^{(t+1)}).$$

Repeating this from a starting value $\theta^{(0)}$ yields a stochastic sequence $\{(\theta^{(t)},Y_{mis}^{(t)})\}$ whose stationary distribution is $P(\theta,Y_{mis}|Y_{obs})$, and the subsequences $\{\theta^{(t)}\}$ and $\{Y_{mis}^{(t)}\}$ have $P(\theta|Y_{obs})$ and $P(Y_{mis}|Y_{obs})$ as their respective stationary distributions. These two steps are also called I-step for imputation and P-step for Posterior. At iteration t, let $Z^{(t)} = (z_1^{(t)},z_2^{(t)},\ldots,z_m^{(t)})$ be a sample of size m from a distribution that approximates the target distribution $P(z)$. This sample is updated in two steps. First, $U^{(t+1)} = (u_1^{(t+1)},u_2^{(t+1)},\ldots,u_m^{(t+1)})$ is created by drawing $u_i^{(t+1)} \sim g(u|v_i^{(t)})$ independently for $i = 1,2,\ldots,m$. Next, $V^{(t+1)} = (v_1^{(t+1)},v_2^{(t+1)},\ldots,v_m^{(t+1)})$ is drawn as an iid sample from the equally weighted mixture of the conditionals $h(v|u_i^{(t+1)})$, $\overline{h}(v|U^{(t+1)}) = m^{-1}\sum_{i=1}^m h(v|u_i^{(t+1)})$, which completes the new sample $Z^{(t+1)}$. From functional analysis, the distribution of $Z^{(t)}$ converges to $P(z)$ as $t \to \infty$.

MCMC is a collection of techniques for creating pseudorandom draws from probability distributions. The goal of he MCMC is to generate one or more values of a random variable Z, which is typically multidimensional. Let $P(Z) = f(Z)$ denote the density of Z, which is called the target distribution. Rather than attempting to draw from f directly, we generate a sequence $\{Z^{(1)},Z^{(2)},\ldots,Z^{(t)},\ldots\}$ where each variate in the sequence depends in some fashion on the preceding ones, and where the stationary distribution is the target f, the samples are used to make inferences about the posterior distribution. Gibbs sampling is the most popular and well known form of MCMC. Suppose that a random vector Z is partitioned into J subvectors $Z = (Z_1,Z_2,\ldots,Z_J)$. Let $P(Z)$ denote the joint distribution of Z. In Gibbs sampling (Geman and Geman 1984), we iteratively draw from the conditional distribution of each subvector given all the others. Given the value of Z at step t, $Z^{(t)} = (Z_1^{(t)},Z_2^{(2)},\ldots,Z_J^{(t)})$, the value of Z at step $t+1$ is obtained by successively drawing from the distributions

$$Z_1^{(t+1)} \sim P(Z_1|Z_2^{(t)},Z_3^{(t)},\ldots,Z_J^{(t)})$$

$$
\begin{aligned}
Z_2^{(t+1)} &\sim P(Z_2|Z_1^{(t+1)}, Z_3^{(t)}, \ldots, Z_J^{(t)}) \\
&\cdots \\
Z_J^{(t+1)} &\sim P(Z_J|Z_2^{(t+1)}, Z_3^{(t+1)}, \ldots, Z_{J-1}^{(t+1)})
\end{aligned}
$$

The sequence $Z^{(t)}$ forms a Markov chain which under mild regularity conditions has a stationary distribution equal to $P(Z)$, i.e., $Z^{(t)} \to Z$ in distribution as $\to \infty$. Gibbs actually refers to a class of probability distribution on lattice systems that have been used in spatial analysis and statistical image reconstruction.

An older method of MCMC is the algorithm of Metropolis et al (1953) and its generalization by Hastings (1970). In the Hastings version, a Markov chain $\{Z^{(t)}\}$ with stationary distribution $P(Z) = f(Z)$ is constructed as follows. Given $Z^{(t)}$, a candidate value $\tilde{Z}$ is drawn from a transition distribution $q(Z; Z^{(t)})$. Then the ratio $R^{(t+1)} = \frac{q(Z^{(t)};\tilde{Z})}{q(\tilde{Z};Z^{(t)})} \frac{f(\tilde{Z})}{f(Z^{(t)})}$ is calculated. If $R^{(t+1)}$ is greater than 1, we accept the value of the candidate variable and set $Z^{(t+1)} = \tilde{Z}$. If $R^{(t+1)} < 1$, we randomly accept the value of $\tilde{Z}$ as our next iterate with probability $R^{(t+1)}$, and otherwise keep the current value. The Metropolis-Hastings algorithms proceeds from the present value Z to the next in three steps.

1. Generate $Z^\star$ from the proposal distribution $q(., Z)$
2. Compute the Hastings ratio A
3. With probability $a = \min(1, A)$ the process moves to $Z^\star$

The acceptance ratio for a move from state Z to $Z^\star$ $A = \frac{p(Z^\star)q(Z;Z^\star)}{p(Z)q(Z^\star;Z)}$ is the product of the probability ratio $p(Z^\star)/p(Z)$ and the ratio of the probability of the proposition of the *reverse move* from $Z^\star$ to Z against the probability of the proposition of the *forward move* from Z to $Z^\star$. Here $p(Z)$ only needs to be known up to a multiplicative constant. If $q(Z^\star; Z) = q(Z; Z^\star)$ then it reduces to Metropolis algorithm. A special case of the Metropolis-Hastings sampler is the Gibbs sampler, with which changes typically are made to one element of Z at a time. Use Z_{-i} to indicate all elements of Z apart from Z_i, then when Z_i is updated, the new value for Z_i is sampled from the conditional distribution $p(Z_i|Z_{-i})$, resulting a acceptance probability of 1 all the time.

$$
A = \frac{p(Z_i^\star, Z_{-i})p(Z_i|Z_{-i})}{p(Z_i, Z_{-i})p(Z_i^\star|Z_{-i})} = \frac{p(Z_i^\star, Z_{-i})}{p(Z_i, Z_{-i})} \times \frac{p(Z_i, Z_{-i})}{p(Z_{-i})} \times \frac{p(Z_{-i})}{p(Z_i^\star, Z_{-i})} = 1
$$

2 Linkage analysis

There were many developments of linkage analysis from 1970s to 1990s, yet the situation with extended family and multiple loci, though more powerful (Wijsman & Amos 1997), remains unsolved. This is further complicated by incorporating ascertainment scheme.

Monte Carlo likelihoods

The likelihood of a pedigree can be written as

$$
L(\theta) = \sum_{\mathbf{x}} p_\theta(Y|X)P_\theta(X) = E_\theta(P_\theta(Y|X))
$$

where X are latent variables. This is an early form given by Ott (1979). Inconsistent realization of X with data Y is discarded so it is rejection sampling. This does not work well except for small pedigrees. An improvement may be some form of importance sampling, $E_h(g(X)) = \sum_{\mathbf{x}} g(X)h(X) = \sum_{\mathbf{x}} g(X)\frac{h(X)}{h(X)^\star}h(X)^\star = E_h^\star\left(g(X)\frac{h(X)}{h(X)^\star}\right)$. So $h(.)^\star$ is simulated and used to reweight the realization. It works best when $h(X)^\star \propto g(X)h(X)$; the tails of the distribution have to be very similar. Again this is not very straightforward. Using the importance sampling idea, the above likelihood can be modified to $L(\theta) = \sum P_\theta(Y|X)\frac{P_\theta(X)}{P_{\theta_0(X)}}P_{\theta_0}(X) = E_{\theta_0}\left(P_\theta(Y|X)\frac{P_\theta(X)}{P_{\theta_0}(X)}\right)$ where only some canonical model θ_0 is simulated.

Normally because genotypes simulated from the prior will bear almost no relation to the observed data, we need to estimate $L(\theta) = P_\theta(Y) = \sum_{\mathbf{x}} P_\theta(Y,X)$, or $\sum_{\mathbf{x}}\left(\frac{P_\theta(Y,X)}{P_{\theta_0(X|Y)}}\right)P_{\theta_0}(X|Y) = E_{\theta_0}\left(\frac{P_\theta(Y,X)}{P_{\theta_0}(X|Y)}|Y\right) = P_{\theta_0}(Y)E_{\theta_0}\left(\frac{P_\theta(Y,X)}{P_{\theta_0}(Y,X)}|Y\right)$. Then a likelihood ratio function is formed and estimated as $\frac{1}{N}\sum_{t=1}^{N}\left(\frac{P_\theta(Y,X^{(t)})}{P_{\theta_0}(Y,X^{(t)})}\right)$.

The reducibility of Gibbs sampler for genetic loci with more than two alleles was first addressed by Sheehan & Thomas (1993). Their method modified either segregation probabilities or the penetrance probabilities and no reweighting was required in order for the realization to represent the distribution of genotypes under the true genetic model. Lin et al. (1993) used similar penetrance modifications to achieve irreducibility but with Metropolis-coupled samplers. Under certain restrictions, updates can be made by sampling from *reduced* conditionals, conditioning on only a subset of X_{-i} (Besag et al. 1995). Reverse jump MCMC (Green 1995; Richardson and Green 1997) is an extension to the Metropolis-Hastings sampler, permitting moves to be made that change the dimension of X. For example, consider a move from X to $X^\star$, where X has dimension l_0 and where $X^\star$ has dimension l_1, with $l_1 > l_0$. To make up the difference in length between X and $X^\star$, a random vector U, of length $l_1 - l_0$, is sampled and then is transformed to yield the extra elements of $X^\star$. When the reverse step is made, the extra elements are simply discarded. The acceptance ratio for this step is given by

$$\frac{p(X^\star)q(l_0;l_1)}{p(X)q(l_1;l_0)q(U)}\left|\frac{\partial X^\star}{\partial(X,U)}\right|$$

where $q(l_1;l_0)$ is the probability of the proposition of the move as described before, $q(U)$ is the proposal probability for the U, and the last term is the Jacobian of the transformation from (X,U) to $X^\star$.

Monte Carlo estimation using meiosis sampler

For all members i of a pedigree at all loci j denote the observed marker and/or trait phenotypes $Y = \{Y_{ij}\}$, we introduce the meiosis indicator variable

$$S_{ij} = \begin{cases} 0 & \text{meiosis } i \text{ locus } j \text{ is maternal} \\ 1 & \text{meiosis } i \text{ locus } j \text{ is paternal} \end{cases}$$

$S_{\star j} = (S_{1j}, \ldots, S_{mj})$ is the inheritance vector for m meiosis. Given L ordered loci define $S_{i\star} = (S_{i1}, \ldots, S_{iL})$ as the vector of meiosis i. The above likelihood can alternatively be written in terms of S_{ij}, noting meioses i are independent:

$$P(Y) = \sum_S P(Y|S)P(S) = \sum_S\left(\prod_j P(Y_{*j}|J(S_{*j}))\right)\left(\prod_i P(S_{i*})\right)$$

where $J(S_{*j})$ is the pattern of gene identity by descent among observed individuals at locus j, and $P(Y_{*j}|J(S_{*j})) = P(Y_{*j}|S_{*j})$.

Further

$$P_\theta(Y) = \sum_S P_\theta(Y,S) = E_{P^\star}\left(\frac{P_\theta(Y,S)}{P^\star(S)}\right)$$

with $P^\star \propto P_\theta(Y,S)$ it will be

$$E_{\theta_0}\left(\frac{P_\theta(Y,S)}{P_{\theta_0}(S|Y)}\middle|Y\right) = P_{\theta_0}(Y)E_{\theta_0}\left(\frac{P_\theta(Y,S)}{P_{\theta_0}(Y|S)}\middle|Y\right)$$

and the likelihood ratio estimate

$$\frac{L(\theta)}{L(\theta_0)} = \frac{P_\theta(Y)}{P_{\theta_0}(Y)} = E_{\theta_0}\left(\frac{P_\theta(Y,S)}{P_{\theta_0}(Y,S)}\middle|Y\right)$$

Now

$$\frac{1}{N}\sum_{t=1}^{N}\left(\frac{P_\theta(Y,S^{(t)})}{P_{\theta_0}(Y,S^{(t)})}\right)$$

is an estimate of $L(\theta)/L(\theta_0)$ in the neighbourhood of θ_0 with $S^{(t)} \sim P_{\theta_0}(S|Y)$.

When we examine trait against map of L markers, $\theta = \{\beta_T, \gamma_T, \Gamma_M\}$ containing the trait segregation model parameters β_T and a hypothesised trait-locus location γ_T, we have,

$$\begin{aligned}\frac{L(\beta_T,\gamma_{T_1},\Gamma_M)}{L(\beta_T,\gamma_{T_0},\Gamma_M)} &= E_{\theta_0}\left(\frac{P_{\theta_1}(Y_T,Y_M|S_T,S_M)P_{\theta_1}(S_T,S_M)}{P_{\theta_0}(Y_T,Y_M|S_T,S_M)P_{\theta_0}(S_T,S_M)}\middle|Y_T,Y_M\right)\\ &= E_{\theta_0}\left(\frac{P_{\theta_{T_1}}(S_T|S_M)}{P_{\theta_{T_0}}(S_T|S_M)}\middle|Y_T,Y_M\right)\end{aligned}$$

Since only the position of of the trait locus differs between numerator and denominator.

It turns out 2-marker is not too bad but the intervals are hard, now the MCMC would come in. The following scheme is desirable since it combines realizations from models in a range of interests, say at θ_j, $j = 0,\ldots,K$

$$P^\star(S) = \frac{1}{\sum_j N_j}\sum_{j=0}^{K} N_j P_{\theta_j}(S|Y) = \frac{1}{\sum_j N_j}\sum_{j=0}^{k} N_j \frac{P_{\theta_j}(Y,S)}{L(\theta_j)}$$

where N_j is sampled from $P_{\theta_j}(S|Y)$ and

$$L(\theta_j) = E_{P^\star}\left(\frac{P_{\theta_j}(Y,S)}{P^\star(S)}\right)$$

and $P^\star$ is chosen to avoid instability. Again we use the approximation

$$L(\theta_j) \approx \sum_{S^\star}\left(\frac{P_{\theta_j}(Y,S^\star)}{\sum_{l=0}^{k} N_l P_{\theta_l}(Y,S^\star)/L(\theta_l)}\right), j = 0,\ldots,K$$

and in fact

$$L(\theta) = \sum_{S^\star} \left(\frac{P_\theta(Y, S^\star)}{\sum_{l=0}^{k} N_l P_{\theta_l}(Y, S^\star)/L(\theta_l)} \right), \forall \theta$$

With these notations the algorithm of Lander & Green (1987) can be introduced.
Let $Y^{(j)} = (Y_{\star 1}, \ldots, Y_{\star j})$, and $R_j(s) = P(Y^{(j)}, S_{\star j} = s)$, so that $Y = Y^{(L)}$,

$$P(Y) = \sum_s R_L(s)$$

and

$$R_j(s) = P(Y_{\star j}|S_{\star j} = s) \sum_{s^\star} P(S_{\star j} = s|S_{\star,j-1} = s^\star) R_{j-1}(s^\star)$$

Fore meiosis sampler (M-sampler) coupled with Baum algorithm (Baum et al. 1970, 1972), we need to compute

$$P(S_{i\star}|\{S_{k\star}, k \neq i\}, Y)$$

Define the cumulative probability for the segregation indicator S_{ij}, given the data at loci up to an including locus j

$$Q_j(s) = P(S_{ij} = s|\{S_{k\star}, k \neq i\}, Y^{(j)}), \quad s = 0, 1$$

Forward (compute):

$$Q_1(s) \propto P(Y_{\star 1}|S_{\star 1})$$

$$Q_j(s) \propto P(Y_{\star j}|S_{\star j})(Q_{j-1}(s)(1 - \theta_{j-1}) + Q_{j-1}(1 - s)\theta_j, j = 2, \ldots, L$$

$$Q_L(s) = P(S_{iL} = s|\{S_{k\star}, k \neq i\}, Y = Y^{(L)})$$

Backward (sample): Suppose S_{ij} has been sampled for $l = j, \ldots, L$ then

$$P(S_{i,j-1} = s|\{S_{k\star}, k \neq i\}, \{S_{il}, l = j, \ldots, L\}, Y)$$

$$\propto Q_{j-1}(s)(|S_{ij} - s|\theta_{j-1} + (1 - |S_{ij} - s|))(1 - \theta_{j-1}))$$

The MCMC procedure for realisation of S from its conditional distribution given data Y as well as the estimation provide location score estimate, together with a direct estimates of posterior probabilities of patterns of gene IBD. Possible extension includes sex-difference in recombination frequencies, interference between tightly linked loci. By now Metropolis-Hastings ratio for interference (I=interference, H=Haldane) is

$$\alpha = \frac{P^{(I)}(S^+, Y)}{P^{(I)}(S, Y)} \frac{P^{(H)}(S_{i*}|S_{l*}, l \neq i, Y)}{P^{(H)}(S_{i*}^+|S_{l*}, l \neq i, Y)}$$

Thompson (1995, 2000a, 2000b, 2000c) gives an example of rare recessive trait in an Iceland extended pedigree with 17 markers.

References

Baum, L. E. (1972). An inequality and associated maximization technique in statistical estimation for probabilistic functions on Markov processes. *Inequalities 3*, 1–8.

Baum, L. E., T. Petrie, G. Soules, and N. Weiss (1970). A maximization techniques occurring in the statistical analysis of probablistic functions of Markov chains. *Ann Math Stat 41*, 164–171.

Besag, J., P. Green, D. Higdon, and K. Mengersen (1995). Bayesian computation and stochastic systems. *Statist Sci 10*, 3–66.

Dempster, A. P., N. Laird, and D. B. Rubin (1977). Maximum likelihood from incomplete data via the EM algorithm. *J Roy Stat Soc B 39*, 1–38.

Geman, S. and D. Geman (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Trans Pat Anal Mach Intel 6*, 721–741.

Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika 82*, 711–732.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika 57*, 97–109.

Heath, S. C. (1997). Markov chain Monte Carlo segregation and linkage analysis for oligogenic models. *Am J Hum Genet 61*, 748–760.

Lin, S. (1996). Monte Carlo methods in genetic analysis. In T. Speed and M. S. Waterman (Eds.), *Genetic Mapping and DNA Sequencing*, The IMA Volumes in Mathematics and its Applications, pp. 15–38. Springer.

Lin, S. (1999). Monte Carlo Bayesian methods for quantitative traits. *Comput Stat Data Anal 31*, 89–108.

Meng, X. L. and D. B. Rubin (1993). Maximum likelihood estimation via the ECM algorithm: A general framework. *Biometrika 80*, 899–909.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and T. E. (1953). Equations of state calculations by fast computing machines. *J Chem Phys 21*, 1087–1091.

Richardson, S. and P. J. Green (1997). On Bayesian analysis of mixtures with an unknown number of components. *J Roy Stat Soc B 59*, 731–792.

Schafer, J. L. (1997). *Analysis of Incomplete Multivariate Data*. Monographs on Statistics and Applied Probability 72. Chapman and Hall.

Sheehan, N. and A. Thomas (1993). On the irreducibility of a Markov chain defined on a space of genotype configurations by a sampling scheme. *Biometrics 49*, 163–175.

Sheehan, N. A. (2000). On the application of Markov chain Monte Carlo methods to genetic analyses on complex pedigrees. *Int Stat Rev 68*, 83–110.

Tanner, M. A. (1996). *Tools for Statistical Inference* (Third ed.). Springer.

Tanner, M. A. and W. H. Wong (1987). The calculation of posterior distributions by data augmentation. *J Am Stat Assoc 82*, 528–540.

Thompson, E. A. (1995). Monte Carlo in genetic analysis. Technical Report 294, University of Washington.

Thompson, E. A. (2000a). MCMC estimation of multi-locus genome sharing and multipoint gene location scores. *Int Stat Rev 68*, 53–73.

Thompson, E. A. (2000b). Monte Carlo methods on genetic structures. In O. E. Barndorff-Nielson, D. R. Cox, and C. Klüppelberg (Eds.), *Complex Stochastic Systems*, Séminaire Européan de Statistique, pp. 179–222. London: Chapman and Hall.

Thompson, E. A. (2000c). *Statistical Inferences from Genetic Data on Pedigrees*, Volume 6 of *NSF-CBMS Regional Conference Series in Probability and Statistics*. Beachwood, Ohio: IMS.

Wijsman, E. M. and C. I. Amos (1997). Genetic analysis of simulated oligogenic traits in nuclear and extended pedigrees: summary of GAW10 contributions. *Genet Epidemiol 14*, 719–735.

Wu, C. F. J. (1983). On the convergence properties of the EM algorithm. *Ann Statist 11*, 95–103.